

I claim:

1. A method for acquiring a knowledge base of associated ideas comprising the steps of:
 - providing a pair of documents representing the same idea in two different languages, wherein the first of said pair of documents is expressed in a first language, and the second of said pair of documents is expressed in a second language;
 - receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;
 - analyzing said first of said pair of documents to identify all occurrences of said query in said first of said pair of documents;
 - selecting a plurality of ranges of words in said second of said pair of documents, wherein said selected ranges correspond to the occurrences of said query in said first of said pair of documents;
 - calculating the frequency of words and word strings contained in said selected ranges;
 - tabulating said frequency based on occurrences of all unique words and word strings from said calculating step; and
 - returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated frequency.

2. The method of claim 1, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges.
- 5 3. A method for acquiring a knowledge base of associated ideas comprising the steps of:

 - providing a plurality of document pairs representing the same idea in two different languages, wherein one set of said plurality of document pairs is expressed in a first language, and a second set of said plurality of document pairs is expressed in a
 - 10 second language;
 - receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;
 - analyzing said first set of said plurality pairs to identify all occurrences of said query in said first set;
 - 15 selecting a plurality of ranges of words in said second set of said plurality pairs, wherein said selected ranges correspond to the occurrences of said query in said first set;
 - calculating the frequency of words and word strings contained in said selected ranges,
 - tabulating said frequency based on occurrences of all unique words and word
 - 20 strings from said calculating step; and
 - returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated frequency.

4. The method of claim 3, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges.

5

5. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:

providing a pair of documents representing the same idea in two different
10 languages, wherein the first of said pair of documents is expressed in a first language, and the second of said pair of documents is expressed in a second language;

receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;

analyzing said first of said pair of documents to identify all occurrences of said
15 query in said first of said pair of documents;

selecting a plurality of ranges of words in said second of said pair of documents,
wherein said selected ranges correspond to the occurrences of said query in said first of
said pair of documents;

calculating the frequency of words and word strings contained in said selected
20 ranges;

tabulating said frequency based on occurrences of all unique words and word
strings from said calculating step; and

returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulating frequency.

5 6. The computer device of claim 5, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges.

7. A computer device including a processor, a memory coupled to the processor, and
10 a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:

providing a plurality of document pairs representing the same idea in two different languages, wherein one set of said plurality of document pairs is expressed in a first language, and a second set of said plurality of document pairs is expressed in a
15 second language;

receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;

analyzing said first set of said plurality pairs to identify all occurrences of said query in said first set;

20 selecting a plurality of ranges of words in said second set of said plurality pairs, wherein said selected ranges correspond to the occurrences of said query in said first set;

calculating the frequency of words and word strings contained in said selected ranges, wherein said frequency is based on occurrences of all unique words and word strings;

- tabulating said frequency based on occurrences of all unique words and word strings from said calculating step; and
- returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated frequency.

10 8. The computer device of claim 7, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges.

9. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

providing a pair of documents representing the same idea in two different languages, wherein the first of said pair of documents is expressed in a first language, and the second of said pair of documents is expressed in a second language;

receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;

analyzing said first of said pair of documents to identify all occurrences of said query in said first of said pair of documents;

selecting a plurality of ranges of words in said second of said pair of documents, wherein said selected ranges correspond to the occurrences of said query in said first of said pair of documents;

5 calculating the frequency of words and word strings contained in said selected ranges;

tabulating said frequency based on occurrences of all unique words and word strings from said calculating step; and

returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated
10 frequency.

10. The computer medium of claim 9, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges.

15

11. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

providing a plurality of document pairs representing the same idea in two different languages, wherein one set of said plurality of document pairs is expressed in a
20 first language, and a second set of said plurality of document pairs is expressed in a second language;

receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;

analyzing said first set of said plurality pairs to identify all occurrences of said query in said first set;

selecting a plurality of ranges of words in said second set of said plurality pairs, wherein said selected ranges correspond to the occurrences of said query in said first set;

5 calculating the frequency of words and word strings contained in said selected ranges,

tabulating said frequency based on occurrences of all unique words and word strings from said calculating step; and

returning a list of occurrences of all unique words and word strings if said unique 10 words and word strings occur in more than one of the selected ranges using said tabulated frequency.

12. The computer medium of claim 11, wherein said calculating step omits the occurrence of a word or word string if the word or word string is a subset of a longer 15 word string that occurs in more than one of the selected ranges.

13. A method to tokenize associations for the efficient transfer of information, comprising the following steps:

creating an association; and
20 tokenizing said association by designating a token to be equal to said association; wherein creating an association includes,

providing a pair of documents representing the same idea in two different languages, wherein the first of said pair of documents is expressed in a first language, and the second of said pair of documents is expressed in a second language,

receiving a query to be analyzed, wherein said query is expressed in said first
5 language, and wherein said query consists of a word or word string;

analyzing said first of said pair of documents to identify all occurrences of said query in said first of said pair of documents,

selecting a plurality of ranges of words in said second of said pair of documents, wherein said selected ranges correspond to the occurrences of said query in said first of
10 said pair of documents,

calculating the frequency of words and word strings contained in said selected ranges omitting the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges,

tabulating said frequency based on occurrences of all unique words and word
15 strings from said calculating step, and

returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated frequency.

20 14. The method of claim 13, further comprising:

transmitting said token from one location to a second location or a plurality of second locations;

analyzing, at said second location or plurality of second locations, said designated token to identify said association; and
providing said association to a user.

- 5 15. A method to tokenize associations for the efficient transfer of information, comprising the following steps:
 creating an association; and
 tokenizing said association by designating a token to be equal to said association;
 wherein creating an association includes,
10 providing a plurality of document pairs representing the same idea in two different languages, wherein one set of said plurality of document pairs is expressed in a first language, and a second set of said plurality of document pairs is expressed in a second language;
 receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word or word string;
15 analyzing said first set of said plurality pairs to identify all occurrences of said query in said first set;
 selecting a plurality of ranges of words in said second set of said plurality pairs, wherein said selected ranges correspond to the occurrences of said query in said first set;
20 calculating the frequency of words and word strings contained in said selected ranges, omitting the occurrence of a word or word string if the word or word string is a subset of a longer word string that occurs in more than one of the selected ranges;

tabulating said frequency based on occurrences of all unique words and word strings from said calculating step; and

returning a list of occurrences of all unique words and word strings if said unique words and word strings occur in more than one of the selected ranges using said tabulated frequency.

16. The method of claim 15, further comprising:

transmitting said token from one location to a second location or a plurality of second locations;

10 analyzing, at said second location or plurality of second locations, said designated token to identify said association; and

providing said association to a user.

17. A method for creating a knowledge base of associated ideas involving a source language, a target language, and a third language, comprising the steps of:

receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;

translating said query into a result expressed in said third language;

translating said result into a second result expressed in said target language; and

20 associating said query with said second result in said target language.

18. A method for creating a knowledge base of associated ideas involving a source language, a target language, and a plurality of third languages, comprising the steps of:

- a. receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;
 - b. translating said query into a result expressed in one of said plurality of third languages;
- 5 c. translating said result into a second result expressed in said target language;
- d. repeating steps b. and c. for each of said plurality of third languages;
 - e. returning each of said second results; and
 - f. associating one or more of said second results and said query for all
- 10 second results produced by two or more of said plurality of languages.

19. The method of claims 17 or 15, including the steps of:
translating said query into a third result in said target language utilizing an existing translation scheme or schemes; returning said third results and adding said
15 returned third results to said returned second results in said target language; and
associating one or more of said second and third results of said query for all second or
third results produced more than once.

20. A computer device including a processor, a memory coupled to the processor, and
20 a program stored in the memory, wherein the computer is configured to execute the
program and perform the steps of:
receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;

translating said query into a result expressed in said third language;
translating said result into a second result expressed in said target language; and
associating said query with said second result in said target language.

5 21. A computer device including a processor, a memory coupled to the processor, and
a program stored in the memory, wherein the computer is configured to execute the
program and perform the steps of:

a. receiving a query to be analyzed, wherein said query is expressed in a
source language, and wherein said query consists of a word or word string;

10 b. translating said query into a result expressed in one of said plurality of
third languages;

c. translating said result into a second result expressed in said target
language;

d. repeating steps b. and c. for each of said plurality of third languages;

15 e. returning each of said second results in said target language; and

f. associating one or more of said second results and said query for all
second results produced by two or more of said plurality of languages.

22. The computer device of claims 20 or 21, further configured to perform the steps
20 of:

translating said query into a third result in said target language utilizing an
existing translation scheme or schemes; returning said third results and adding said
returned third results to said returned second results in said target language; and

associating said query to one or more of said second and third results of said query for all second or third results produced more than once.

23. A computer readable storage medium having stored thereon a program executable

5 by a computer processor for performing the steps of:

receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;

translating said query into a result expressed in said third language;

translating said result into a second result expressed in said target language; and

10 associating said query with said second result in said target language.

24. A computer readable storage medium having stored thereon a program executable

by a computer processor for performing the steps of:

a. receiving a query to be analyzed, wherein said query is expressed in a

15 source language, and wherein said query consists of a word or word string;

b. translating said query into a result expressed in one of said plurality of third languages;

c. translating said result into a second result expressed in said target language;

20 d. repeating steps b. and c. for each of said plurality of third languages;

e. returning each of said second results; and

f. associating one or more of said second results and said query for all second results produced by two or more of said plurality of languages.

25. The computer medium of claims 23 or 24, further performing the steps of translating said query into a third result in said target language utilizing an existing translation scheme or schemes; returning said third results and adding said returned third results to said returned second results in said target language; and associating said query to one or more of said second and third results of said query for all second or third results produced more than once.

26. A method to tokenize associations for the efficient transfer of information, comprising the following steps:

creating an association involving a source language, a target language, and a third language, using the following steps:

receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;

translating said query into a result expressed in said third language;

translating said result into a second result expressed in said target language;

associating said query with said second result in said target language; and

tokenizing said association by designating a token to be equal to said association.

20 27. The method of claim 26, further comprising:

transmitting said token from one location to a second location or a plurality of second locations;

analyzing, at said second location or plurality of second locations, said designated token to identify said association; and
providing said association to a user.

5 28. A method to tokenize associations for the efficient transfer of information, comprising the following steps:

creating an association involving a source language, a target language, and a plurality of third languages, using the following steps:

10 a. receiving a query to be analyzed, wherein said query is expressed in a source language, and wherein said query consists of a word or word string;

b. translating said query into a result expressed in one of said plurality of third languages;

c. translating said result into a second result expressed in said target language;

15 d. repeating steps b. and c. for each of said plurality of third languages;

e. returning each of said second results;

f. associating one or more of said second results and said query for all second results produced by two or more of said plurality of languages; and

tokenizing said association by designating a token to be equal to said association.

20

29. The method of claim 28, further comprising:

transmitting said token from one location to a second location or a plurality of second locations;

analyzing, at said second location or plurality of second locations, said designated token to identify said association; and
providing said association to a user.

5 30. A method for creating a knowledge base of associated ideas comprising the steps of:

providing a translation of words expressed in a first language to words and/or word strings expressed in a second language;

10 providing a corpus of documents expressed in said second language;
receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word string;

identifying for said query, all translations of each word comprising said word string query, to said second language utilizing said provided translation;

analyzing said corpus of documents for word strings expressed in said second 15 language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in a first language in said identifying step, wherein said analyzing only counts one translation for each of said words expressed in a first language; and

20 returning a list of said word strings expressed in said second language from said analysis of said corpus of documents as word string results.

31. The method of claim 30, wherein said word strings expressed in said second language have at least a first portion and a second portion, and wherein said list represents associations of said query in said first language to expressions in said second language, further comprising the steps of:

5 examining said list of returned word string results for occurrences wherein any two said returned word string results have overlapping said first and second portions;

combining all of said two overlapping returned word strings into third word strings, wherein said third word strings are a combination of said first word strings and said second word strings, merging said overlapped words; and

10 adding all said third word strings to said list of said word string results.

32. A method of claim 30 where a word expressed in a first language includes certain word strings in a first language such as idioms and collocations.

15 33. The method of claims 30, 31, and 32, further comprising:

ranking said list of word string results based on user-defined criteria.

34. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the 20 program and perform the steps of:

providing a translation of words expressed in a first language to words and/or word strings expressed in a second language;

providing a corpus of documents expressed in said second language;

receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word string;

identifying for said query, all translations of each word comprising said word string query, to said second language utilizing said provided translation;

5 analyzing said corpus of documents for word strings expressed in said second language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in a first language in said identifying step, wherein said analyzing only counts one

10 translation for each of said words expressed in a first language; and

 returning a list of said word strings expressed in said second language from said analysis of said corpus of documents as word string results.

35. The method of claim 34, wherein said word strings expressed in said second language have at least a first portion and a second portion, and wherein said list represents associations of said query in said first language to expressions in said second language, further configured to execute the steps of:

 examining said list of returned word string results for occurrences wherein any two said returned word string results have overlapping said first and second portions;

20 combining all of said two overlapping returned word strings into third word strings, wherein said third word strings are a combination of said first word strings and said second word strings, merging said overlapped words; and

 adding all said third word strings to said list of said word string results.

36. The computer device of claim 34, wherein a word expressed in a first language includes word strings in a first language such as idioms and collocations.

5 37. The computer device of claim 34, further configured to perform the step of ranking said list of word string results based on user-defined criteria.

38. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

10 providing a translation of words expressed in a first language to words and/or word strings expressed in a second language;

providing a corpus of documents expressed in said second language;
receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word string;

15 identifying for said query, all translations of each word comprising said word string query, to said second language utilizing said provided translation;

analyzing said corpus of documents for word strings expressed in said second language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings

20 having translations obtained from a user defined minimum number of words expressed in a first language in said identifying step, wherein said analyzing only counts one translation for each of said words expressed in a first language; and

returning a list of said word strings expressed in said second language from said analysis of said corpus of documents as word string results.

39. The computer medium of claim 38, wherein said word strings expressed in said
5 second language have at least a first portion and a second portion, and wherein said list
represents associations of said query in said first language to expressions in said second
language, further performing the step of:

examining said list of returned word string results for occurrences wherein any
two said returned word string results have overlapping said first and second portions;

10 combining all of said two overlapping returned word strings into third word
strings, wherein said third word strings are a combination of said first word strings and
said second word strings, merging said overlapped words; and

adding all said third word strings to said list of said word string results.

15 40. The computer medium of claim 38, wherein a word expressed in a first language
includes word strings in a first language such as idioms and collocations.

41. The computer medium of claim 38, further performing the step of ranking said list
of word string results based on user-defined criteria.

20 42. A method to tokenize associations for the efficient transfer of information,
comprising the following steps:

creating an association; and

tokenizing said association by designating a token to be equal to said association;
wherein creating an association includes,
providing a translation of words expressed in a first language to words and/or
word strings expressed in a second language;
5 providing a corpus of documents expressed in said second language;
receiving a query to be analyzed, wherein said query is expressed in said first
language, and wherein said query consists of a word string;
identifying for said query, all translations of each word comprising said word
string query, to said second language utilizing said provided translation;
10 analyzing said corpus of documents for word strings expressed in said second
language, wherein said analysis only identifies word strings having a user defined
maximum number of words, and wherein said analysis only identifies word strings
having translations obtained from a user defined minimum number of words expressed in
a first language in said identifying step, wherein said analyzing only counts one
15 translation for each of said words expressed in a first language;
returning a list of said word strings expressed in said second language from said
analysis of said corpus of documents as a result.

43. The method of claim 42, further comprising:

20 transmitting said token from one location to a second location or a plurality of
second locations;
analyzing, at said second location or plurality of second locations, said designated
token to identify said association; and

providing said association to a user.

44. The method of claim 42, wherein a word expressed in a first language includes word strings in a first language such as idioms and collocations.

5

45. The method of claim 30, further comprising:

providing a corpus of documents expressed in said first language;

identifying a user defined number of occurrences of said query in said corpus of documents expressed in said first language;

10 analyzing a user defined number of words and/or word strings to the left and to the right of each of said occurrences of said query and identifying word strings comprising the user defined number of words and/or word strings to the left of said query, said query, and the user defined number of words and/or word strings to the right of said query;

15 creating a list of returned word strings comprising the results of said analyzing step;

analyzing each returned word string individually and identifying all translations of each word comprising each of said returned word strings, to said second language utilizing said provided translation;

20 analyzing said corpus of documents for word strings expressed in said second language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in

the word string in a first language determined by said creating step, wherein said analyzing said corpus counts only one translation for each of said words expressed in said first language;

- returning a list of said second word strings expressed in said second language
- 5 from said analysis of said corpus of documents as a result;
 - analyzing said list of word strings and said list of second word strings to identify the number of occurrences wherein each word string on said list of word strings occurs as a word string subset of a word string on said list of second word strings; and
 - returning a list based on said analyzing said list of word strings and said list of
- 10 second word strings step.

- 46. The method of claim 45, wherein said analyzing said list of word strings and said list of second words strings step includes modifying said number of occurrences by omitting each occurrence of a word string if the word string is a subset of a longer word
- 15 string that is also on the returned list.

- 47. The method of claim 45, wherein a word expressed in a first language includes word strings in a first language such as idioms and collocations.

- 20 48. The method of claim 45 or 46, further comprising:
 - ranking said list of word string results based on user-defined criteria.

- 49. The computer device of claim 34, further configured to perform the steps of:

- providing a corpus of documents expressed in said first language;
- identifying a user defined number of occurrences of said query in said corpus of documents expressed in said first language;
- analyzing a user defined number of words and/or word strings to the left and to
- 5 the right of each of said occurrences of said query and identifying word strings comprising the user defined number of words and/or word strings to the left of said query, said query, and the user defined number of words and/or word strings to the right of said query;
- creating a list of returned word strings comprising the results of said analyzing
- 10 step;
- analyzing each returned word string individually and identifying all translations of each word comprising each of said returned word strings, to said second language utilizing said provided translation;
- analyzing said corpus of documents for word strings expressed in said second
- 15 language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in the word string in a first language determined by said creating step, wherein said analyzing said corpus counts only one translation for each of said words expressed in said
- 20 first language;
- returning a list of said second word strings expressed in said second language from said analysis of said corpus of documents as a result;

analyzing said list of word strings and said list of second word strings to identify the number of occurrences wherein each word string on said list of word strings occurs as a word string subset of a word string on said list of second word strings;

returning a list based on said analyzing said list of word strings and said

5 list of second word strings step.

50. The computer device of claim 49, wherein said analyzing said list of word strings and said list of second words strings step includes modifying said number of occurrences by omitting each occurrence of a word string if the word string is a subset of a longer

10 word string that is also on the returned list.

51. The computer device of claim 49, wherein a word expressed in a first language includes word strings in a first language such as idioms and collocations.

15 52. The computer device of claim 49 or 50, further configured to perform the step of ranking said list of word string results based on user-defined criteria.

53. The computer readable storage medium claim 38, further configured to perform the steps of:

20 providing a corpus of documents expressed in said first language;
 identifying a user defined number of occurrences of said query in said corpus of documents expressed in said first language;

analyzing a user defined number of words and/or word strings to the left and to
the right of each of said occurrences of said query and identifying word strings
comprising the user defined number of words and/or word strings to the left of said
query, said query, and the user defined number of words and/or word strings to the right
5 of said query;

 creating a list of returned word strings comprising the results of said analyzing
step;

 analyzing each returned word string individually and identifying all translations of
each word comprising each of said returned word strings, to said second language
10 utilizing said provided translation;

 analyzing said corpus of documents for word strings expressed in said second
language, wherein said analysis only identifies word strings having a user defined
maximum number of words, and wherein said analysis only identifies word strings
having translations obtained from a user defined minimum number of words expressed in
15 each word string in a first language determined by said creating step, wherein said
analyzing said corpus counts only one translation for each of said words expressed in said
first language;

 returning a list of said second word strings expressed in said second language
from said analysis of said corpus of documents as a result;

20 analyzing said list of word strings and said list of second word strings to identify
the number of occurrences wherein each word string on said list of word strings occurs as
a word string subset of a word string on said list of second word strings; and

returning a list based on said analyzing said list of word strings and said list of second word strings step.

54. The computer medium of claim 53, wherein said analyzing said list of word
5 strings and said list of second words strings step includes modifying said number of occurrences by omitting each occurrence of a word string if the word string is a subset of a longer word string that is also on the returned list.

55. The computer medium of claim 53, wherein a word expressed in a first language
10 includes word strings in a first language such as idioms and collocations.

56. The computer medium of claim 53, further performing the step of ranking said list of word string results based on user-defined criteria.

15 57. A method to tokenize associations for the efficient transfer of information, comprising the following steps:

creating an association; and

tokenizing said association by designating a token to be equal to said association;

wherein creating an association includes,

20 providing a translation of words expressed in a first language to words and/or word strings expressed in a second language;

providing a corpus of documents expressed in said second language;

- receiving a query to be analyzed, wherein said query is expressed in said first language, and wherein said query consists of a word string;
- identifying for said query, all translations of each word comprising said word string query, to said second language utilizing said provided translation;
- 5 analyzing said corpus of documents for word strings expressed in said second language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in a first language in said identifying step, wherein said analyzing only counts one
- 10 translation for each of said words expressed in a first language;
- returning a list of said word strings expressed in said second language from said analysis of said corpus of documents as a result;
- providing a corpus of documents expressed in said first language;
- identifying a user defined number of occurrences of said query in said corpus of
- 15 documents expressed in said first language;
- analyzing a user defined number of words and/or word strings to the left and to the right of each of said occurrences of said query and identifying word strings comprising the user defined number of words and/or word strings to the left of said query, said query, and the user defined number of words and/or word strings to the right
- 20 of said query;
- creating a list of returned word strings comprising the results of said analyzing step;

analyzing each returned word string individually and identifying all translations of each word comprising each of said returned word strings, to said second language utilizing said provided translation;

analyzing said corpus of documents for word strings expressed in said second
5 language, wherein said analysis only identifies word strings having a user defined maximum number of words, and wherein said analysis only identifies word strings having translations obtained from a user defined minimum number of words expressed in the word string in a first language determined by said creating step, wherein said analyzing said corpus counts only one translation for each of said words expressed in said
10 first language;

returning a list of said second word strings expressed in said second language from said analysis of said corpus of documents as a result;

analyzing said list of word strings and said list of second word strings to identify the number of occurrences wherein each word string on said list of word strings occurs as
15 a word string subset of a word string on said list of second word strings;

returning a list based on said analyzing said list of word strings and said list of second word strings step.

58. The method of claim 57, further comprising:

20 transmitting said token from one location to a second location or a plurality of second locations;

analyzing, at said second location or plurality of second locations, said designated token to identify said association; and

providing said association to a user.

59. The method of claim 57, wherein a word expressed in a first language includes word strings in a first language such as idioms and collocations.

5

60. A method for acquiring a knowledge base of associated ideas comprising the steps of:

providing a translation of word strings expressed in a source language to word strings expressed in a target language;

- 10 receiving two segments of content expressed in said source language, wherein said first segment and said second segment have overlapping portions of said content;

translating, using said translation of word strings, said first segment of content to return a third segment expressed in said target language;

- 15 translating, using said translation of word strings, said second segment of content to return a fourth segment expressed in said target language;

analyzing said third segment and said fourth segment to determine if said third segment and said fourth segment have overlapping portions;

associating, if said third segment and said fourth segment have overlapping portions, the overlapping portions of said third segment and said fourth segment with the 20 overlapping portions of said first segment and said second segment; and

associating, if said third segment and said fourth segment have overlapping portions, the combination of said third segment and said fourth segment as a single target language word string, merging said overlapping portions, with the combination of said

first segment and said second segment as a single source word string, merging said overlapping portions.

61. A computer device including a processor, a memory coupled to the processor, and
5 a program stored in the memory, wherein the computer is configured to execute the
program and perform the steps of:

providing a translation of word strings expressed in a source language to word
strings expressed in a target language;

receiving two segments of content expressed in said source language, wherein
10 said first segment and said second segment have overlapping portions of said content;
translating, using said translation of word strings, said first segment of content to
return a third segment expressed in said target language;

translating, using said translation of word strings, said second segment of content
to return a fourth segment expressed in said target language;

15 analyzing said third segment and said fourth segment to determine if said third
segment and said fourth segment have overlapping portions;

associating, if said third segment and said fourth segment have overlapping
portions, the overlapping portions of said third segment and said fourth segment with the
overlapping portions of said first segment and said second segment; and

20 associating, if said third segment and said fourth segment have overlapping
portions, the combination of said third segment and said fourth segment as a single target
language word string, merging said overlapping portions, with the combination of said

first segment and said second segment as a single source word string, merging said overlapping portions.

62. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:
 - providing a translation of word strings expressed in a source language to word strings expressed in a target language;
 - receiving two segments of content expressed in said source language, wherein said first segment and said second segment have overlapping portions of said content;
 - 10 translating, using said translation of word strings, said first segment of content to return a third segment expressed in said target language;
 - translating, using said translation of word strings, said second segment of content to return a fourth segment expressed in said target language;
 - analyzing said third segment and said fourth segment to determine if said third
 - 15 segment and said fourth segment have overlapping portions;
 - associating, if said third segment and said fourth segment have overlapping portions, the overlapping portions of said third segment and said fourth segment with the overlapping portions of said first segment and said second segment; and
 - associating, if said third segment and said fourth segment have overlapping
 - 20 portions, the combination of said third segment and said fourth segment as a single target language word string, merging said overlapping portions, with the combination of said first segment and said second segment as a single source word string, merging said overlapping portions.

63. A method to tokenize associations for the efficient transfer of information, comprising the following steps:
- creating an association; and
 - 5 tokenizing said association by designating a token to be equal to said association; wherein creating an association includes,
 - providing a translation of word strings expressed in a source language to word strings expressed in a target language;
 - receiving two segments of content expressed in said source language, wherein 10 said first segment and said second segment have overlapping portions of said content;
 - translating, using said translation of word strings, said first segment of content to return a third segment expressed in said target language;
 - translating, using said translation of word strings, said second segment of content to return a fourth segment expressed in said target language;
 - 15 analyzing said third segment and said fourth segment to determine if said third segment and said fourth segment have overlapping portions;
 - associating, if said third segment and said fourth segment have overlapping portions, the overlapping portions of said third segment and said fourth segment with the overlapping portions of said first segment and said second segment;
 - 20 associating, if said third segment and said fourth segment have overlapping portions, the combination of said third segment and said fourth segment as a single target language word string, merging said overlapping portions, with the combination of said

first segment and said second segment as a single source word string, merging said overlapping portions.

64. The method of claim 63, further comprising:

5 transmitting said token from one location to a second location or a plurality of second locations;

analyzing, at said second location or plurality of second locations, said designated token to identify said association; and

providing said association to a user.

10

65. A method for converting content and reconstructing a knowledge base comprising the steps of:

a. receiving content expressed in a first language;

b. parsing said content expressed in a first language into a plurality of

15 segments;

c. selecting a first segment and a second segment, with said first segment

having an overlapping portion of said content with said second segment;

d. accessing a first target segment of said content expressed in a second

language, said first target segment corresponding to one of said first and second

20 segments;

e. accessing a second target segment of said content expressed in the second

language, said second target segment corresponding to the other one of said first and

second segments and having an overlapping portion with said first target segment;

- f. determining said content expressed in the second language based on combining said first target and second target segments, merging overlapping portions;
 - g. providing said content expressed in said second language; and
 - h. repeating steps c. through g. for all of said plurality of segments, wherein
- 5 the second segment is designated as the first segment, and a next segment, with overlapping portions with the second segment, is designated as the second segment; and
- i. repeating step h. for all next segments in said plurality of segments until all of said content is converted into said second language.

10 66. A method for converting content of a document by reconstructing a knowledge base comprising the steps of utilizing a database of segment associations between content in a first language and a second language wherein said conversion includes parsing and examining overlapping segments of content of the document in said first language with their respective translations that have overlapping segments of content in said second

15 language, and merging overlapping segments from said examined first language content and said examined second language content, and associating the content of said first language content with said second language content after merging overlapping segments.

67. A method of converting a document and reconstructing a knowledge base, the

20 method comprising the steps of:

- a. providing content comprising data segments in a first language associated with data segments in a second language;
- b. selecting from the document to be translated in a first language a data segment that begins with the first word of the document and exists in a database;

- c. retrieving from the database a segment in the second language associated with the located first segment in the first language;
- d. selecting at least a second delimited segment in the first language that has one or more overlapping portions with the previous delimited segment in the first language;
- 5 e. retrieving from the database a second segment in the second language associated with the selected second segment in the first language;
- f. returning the two data segments in the first language and merging the overlapping portions as a single data segment in the first language;
- 10 g. returning, if the two data segments in the second language have overlapping portions, a single data segment in the second language merging the overlapping portions; and
- 15 h. associating said single data segment in said first language with said single data segment in said second language, thereby returning a conversion of said single data segment from said first language to said second language.

68. The method of claim 67, further comprising repeating steps d. through h. designating a next data segment in the first language document that overlaps with the prior data segment in a first language as a second delimited segment in the first language.

20

69. The method of claim 68, further comprising repeating steps d. through h. for all next data segments of the first language document that overlap with the prior data segment in the first language until the entire document is converted.

70. The method of claim 67, wherein said segments occur in the form of a word or a plurality of words.

5 71. The method of claim 67, wherein said segments occur in the form of a plurality of words.

72. A method of converting a document, the method comprising the steps of:

a. providing content comprising data segments in a first language associated

10 with data segments in a second language;

b. selecting from the document to be translated in a first language a data

segment that begins with the first word of the document and exists in a database;

c. retrieving from the database a segment in the second language associated
with the located first segment in the first language;

15 d. selecting at least a second delimited segment in the first language that has
one or more overlapping portions with the previous delimited segment in the first
language;

e. retrieving from the database a second segment in the second language
associated with the selected second segment in the first language that has an overlapping
20 portion with the segment in the second language;

f. combining the two segments in the second language, merging the
overlapping portions, to form a translation of the two segments in the first language,
merging overlapping portions.

73. The method of claim 72, further comprising repeating steps d. through f.

designating a next segment as a second delimited segment until the document is completely converted into a second language.

5

74. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:

a. receiving content expressed in a first language;

10 b. parsing said content expressed in a first language into a plurality of segments;

c. selecting a first segment and a second segment, with said first segment having an overlapping portion of said content with said second segment;

15 d. accessing a first target segment of said content expressed in a second language, said first target segment corresponding to one of said first and second segments;

e. accessing a second target segment of said content expressed in the second language, said second target segment corresponding to the other one of said first and second segments and having an overlapping portion with said first target segment;

20 f. determining said content expressed in the second language based on combining said first target and second target segments, merging overlapping portions;

g. providing said content expressed in said second language; and

- h. repeating steps c. through g. for all of said plurality of segments, wherein the second segment is designated as the first segment, and a next segment, with overlapping portions with the second segment, is designated as the second segment; and
 - i. repeating step h. for all next segments in said plurality of segments until 5 all of said content is converted into a second language.

75. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:
- 10 a. providing content comprising data segments in a first language associated with data segments in a second language;
 - b. selecting from the document to be translated in a first language a data segment that begins with the first word of the document and exists in a database;
 - c. retrieving from the database a segment in the second language associated 15 with the located first segment in the first language;
 - d. selecting at least a second delimited segment in the first language that has one or more overlapping portions with the previous delimited segment in the first language;
 - e. retrieving from the database a segment in the second language associated 20 with the selected second segment in the first language;
 - f. returning the two data segments in the first language and merging the overlapping portions as a single data segment in the first language;

g. returning, if the two data segments in the second language have overlapping portions, a single data segment in the second language combining the overlapping portions; and

h. associating said single data segment in said first language with said single
5 data segment in said second language, thereby returning a conversion of said single data segment from said first language to said second language.

76. The computer device of claim 75, further configured to repeat steps d. through h.
designating a next data segment in the first language document that overlaps with the
10 prior data segment in a first language as a second delimited segment in the first language.

77. The computer device of claim 76, further comprising repeating steps d. through h.
for all next data segments of the first language document that overlap with the prior data
segment in the first language until the content of the entire document is converted.

15

78. The computer device of claim 75, wherein said segments occur in the form of a
word or a plurality of words.

79. The computer device of claim 75, wherein said segment occur in the form of a
20 plurality of words.

80. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:
- a. providing content comprising data segments in a first language associated with data segments in a second language;
 - 5 b. selecting from the document to be translated in a first language a data segment that begins with the first word of the document and exists in a database;
 - c. retrieving from the database a segment in the second language associated with the located first segment in the first language;
 - 10 d. selecting at least a second delimited segment in the first language that has one or more overlapping portions with the previous delimited segment in the first language;
 - e. retrieving from the database a second segment in the second language associated with the selected second segment in the first language that has an overlapping portion with the segment in the second language;
 - 15 f. combining the two segments in the second language, merging the overlapping portions, to form a translation of the two segments in the first language, merging overlapping portions.
- 20 81. The computer device of claim 80, further configured to repeat steps d. through f. designating a next segment as a second delimited segment until the document is completely converted into a second language.

82. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:
- a. receiving content expressed in a first language;
 - b. parsing said content expressed in a first language into a plurality of segments;
 - c. selecting a first segment and a second segment, with said first segment having overlapping portions of said content with said second segment;
 - d. accessing a first target segment of said content expressed in a second language, said first target segment corresponding to one of said first and second segments;
 - e. accessing a second target segment of said content expressed in the second language, said second target segment corresponding to the other one of said first and second segments and having an overlapping portion with said first target segment;
 - f. determining said content expressed in the second language based on combining said first target and second target segments, merging overlapping portions;
 - g. providing said content expressed in said second language; and
 - h. repeating steps c. through g. for all of said plurality of segments, wherein the second segment is designated as the first segment, and a next segment, with overlapping portions with the second segment, is designated as the second segment; and
- i. repeating step h. for all next segments in said plurality of segments.

83. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

- a. providing content comprising data segments in a first language associated with data segments in a second language;
- b. selecting from the document to be translated in a first language a data segment that begins with the first word of the document and exists in a database;
- 5 c. retrieving from the database a segment in the second language associated with the located first segment in the first language;
- d. selecting at least a second delimited segment in the first language that has one or more overlapping portions with the previous delimited segment in the first language;
- 10 e. retrieving from the database a second segment in the second language associated with the selected second segment in the first language;
- f. returning the two data segments in the first language and merging the overlapping portions as a single data segment in the first language;
- 15 g. returning, if the two data segments in the second language have overlapping portions, a single data segment in the second language combining the overlapping portions; and
- h. associating said single data segment in said first language with said single data segment in said second language, thereby returning a conversion of said single data segment from said first language to said second language.

20

84. The computer medium of claim 83, further configured to repeat steps d. through h. designating a next data segment in the first language document that overlaps with the prior data segment in a first language as a second delimited segment in the first language.

85. The computer medium of claim 84, further comprising repeating steps d. through h. for all next data segments of the first language document that overlap with the prior data segment in the first language until the content of the entire document is converted.

5

86. The computer medium of claim 84, wherein said segments occur in the form of a word or a plurality of words.

87. The computer medium of claim 83, wherein said segments occur in the form of a 10 plurality of words.

88. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

- a. providing content comprising data segments in a first language associated 15 with data segments in a second language;
- b. selecting from the document to be translated in a first language a data segment that begins with the first word of the document and exists in a database;
- c. retrieving from the database a segment in the second language associated with the located first segment in the first language;
- d. selecting at least a second delimited segment in the first language that has 20 one or more overlapping portions with the previous delimited segment in the first language;

- e. retrieving from the database a second segment in the second language associated with the selected second segment in the first language that has an overlapping portion with the segment in the second language;
- f. combining the two segments in the second language, merging the overlapping portions, to form a translation of the two segments in the first language, merging overlapping portions.

5 89. The computer medium of claim 88, further configured to repeat steps d. through f. designating a next segment with overlapping portions as a second delimited segment in
10 the first language until the document is completely converted into a second language.

90. A computer system for converting content and reconstructing a knowledge base, comprising:

- a. a computing device that receives content expressed in a first language and parses said content into at least a first segment and a second segment, said first segment having a first portion, said second segment having a second portion, said first portion and said second portion having overlapping portions of said content;

15 b. wherein said computing device accesses third and fourth segments of said content that are each expressed in a second language, said third segment corresponding to one of said first and second segments, said fourth segment corresponding to the other of
20 said first and second segments and having an overlapping portion with said third segment; and

c. wherein said computing device determines said content expressed in the second language based on said third and fourth segments having an overlapping portion and provides said content in the second language.

5 91. The computer system defined in claim 90, further comprising a database system which stores said third and fourth segments, wherein said computing device accesses said third and fourth segments from said database system.

92. The computer system defined in claim 90, wherein said second segment of
10 content is designated as the first segment of content in a first language, and a next segment of content in a first language that has an overlapping portion with the designated first segment in a first language is designated as the second segment of content in a first language and repeating steps a. through c. for each next segment of content until the entire content is converted.

15

93. A method for creating a frequency association database in a single language comprising:

providing a collection of documents, wherein said collection includes at least one document;

20 receiving from a user a word or word string query to be analyzed;
searching said collection of documents for occurrences of said query;
creating a list of words and word strings occurring within a user-defined amount of words of said query; and

tabulating a list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said query.

94. The method of claim 93, further comprising the steps of creating a list of the
5 proximity of said words and word strings occurring within a user-defined amount of words of said query.

95. The method of claim 93, further comprising associating two or more words or word strings or both on said list of words.

10

96. The method of claims 93 or 94, wherein one or more of said list of words and word strings, said list of frequency of occurrences, and said list of the proximity of said words and word strings is returned to a user.

15 97. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:

providing a collection of documents, wherein said collection includes at least one document;

20 receiving from a user a word or word string query to be analyzed;
searching said collection of documents for occurrences of said query;
creating a list of words and word strings occurring within a user-defined amount of words of said query; and

tabulating a list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said query.

98. The computer device of claim 97, further configured to create a list of the
5 proximity of said words and word strings occurring within a user-defined amount of words of said query.

99. The computer device of claim 97, further comprising associating two or more words or word strings or both on said list of words.

10

100. The computer device of claim 97 or 98, wherein one or more of said list of words and word strings, said list of frequency of occurrences, and said list of the proximity of said words and word strings is returned to a user.

15 101. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

providing a collection of documents, wherein said collection includes at least one document;

receiving from a user a word or word string query to be analyzed;

20 searching said collection of documents for occurrences of said query;

creating a list of words and word strings occurring within a user-defined amount of words of said query; and

tabulating a list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said query.

102. The computer medium of claim 101, further performing the steps of creating a list
5 of the proximity of said words and word strings occurring within a user-defined amount of words of said query.

103. The computer medium of claim 101, further comprising associating two or more words or word strings or both on said list of words.

10

104. The computer medium of claims 101 or 102, wherein one or more of said list of words and word strings, said list of frequency of occurrences, and said list of the proximity of said words and word strings is returned to a user.

15 105. The method of claim 93, further comprising:

receiving from a user a second word or word string query to be analyzed;
searching said collection of documents for occurrences of said second query;
creating a second list of words and word strings occurring within a user-defined amount of words of said second query;

20 creating a second list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said second query;
creating a third list of words and word strings that occur on both of said list of words and word strings and said second list of words and word strings within a user

defined number of words of the query and a user defined number of words of the second query; and

associating words and word strings on said third list with said first query and said second query.

5

106. The method of claim 105, wherein said third list of words and word strings is modified by user-defined criteria.

107. The method of claim 105, wherein said third list of said words and word strings is

ranked based on user-defined parameters.

108. The computer device of claim 97, further configured to perform the steps of:

receiving from a user a second word or word string query to be analyzed;

searching said collection of documents for occurrences of said second query;

15 creating a second list of words and word strings occurring within a user-defined

amount of words of said second query;

creating a second list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said second query;

creating a third list of words and word strings that occur on both of said list of

20 words and word strings and said second list of words and word strings within a user

defined number of words of the query and a user defined number of words of the second

query; and

associating words and word strings on said third list with said first query and said second query.

109. The computer device of claim 108, wherein said third list of words and word
5 strings is modified by user-defined criteria.

110. The computer device of claim 108, wherein said third list of words and word strings is ranked based on user-defined parameters.

10 111 The computer medium of claim 101, further comprising:
receiving from a user a second word or word string query to be analyzed;
searching said collection of documents for occurrences of said second query;
creating a second list of words and word strings occurring within a user-defined amount of words of said second query;
15 creating a second list of frequency of occurrences of all recurring words and word strings occurring within a user-defined amount of words of said second query;
creating a third list of words and word strings that occur on both of said list of words and word strings and said second list of words and word strings within a user defined number of words of the query and a user defined number of words of the second
20 query; and
associating words and word strings on said third list with said first query and said second query.

112. The computer medium of claim 111, wherein said third list of words and word strings is modified by user-defined criteria.

113. The computer medium of claim 111, wherein said third list of words and word strings is ranked based on user-defined parameters.

114. A method for associating words in a language comprising:

providing a collection of documents; wherein said collection includes at least one document;

10 selecting a first word or word string, and a second word or word string;

locating all documents having occurrences of the first word or word string within a defined proximity range of the second word or word string, with said defined proximity range having an upper limit and a lower limit;

defining in the located documents a range, wherein the range is defined in relation 15 to the first word or word string and the second word or word string;

searching said ranges for recurring words and word strings; and
associating the first word or word string and the second word or word string with recurring words and word strings based on frequency of occurrence of the recurring words and word strings within the ranges.

20

115. The method of claim 114, wherein said associating first word or word string and second word or word string is enhanced by a greater frequency of occurrence of a word or word string.

116. The method of claim 114, wherein said associating first word or word string and second word or word string is enhanced by a lesser frequency of occurrence of a word or word string.

5

117. The method of claim 114, wherein said upper and said lower limit of said defined proximity range are equal.

118. A computer device including a processor, a memory coupled to the processor, and
10 a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:

providing a collection of documents; wherein said collection includes at least one document;

selecting a first word or word string, and a second word or word string;

15 locating all documents having occurrences of the first word or word string within a defined proximity range of the second word or word string, with said defined proximity range having an upper limit and a lower limit;

defining in the located documents a range, wherein the range is defined in relation to the first word or word string and the second word or word string;

20 searching said ranges for recurring words and word strings; and

associating the first word or word string and the second word or word string with recurring words and word strings based on frequency of occurrence of the recurring words and word strings within the ranges.

119. The computer device of claim 118, wherein said associating first word or word string and second word or word string is enhanced by a greater frequency of occurrence of a word or word string.

5

120. The computer device of claim 118, wherein said associating first word or word string and second word or word string is enhanced by a lesser frequency of occurrence of a word or word string.

10 121. The computer device of claim 118, wherein said upper and said lower limit of said defined proximity range are equal.

122. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:

15 providing a collection of documents; wherein said collection includes at least one document;

selecting a first word or word string, and a second word or word string;

locating all documents having occurrences of the first word or word string within a defined proximity range of the second word or word string, with said defined proximity

20 range having an upper limit and a lower limit;

defining in the located documents a range, wherein the range is defined in relation to the first word or word string and the second word or word string;

searching said ranges for recurring words and word strings; and

associating the first word or word string and the second word or word string with recurring words and word strings based on frequency of occurrence of the recurring words and word strings within the ranges.

5 123. The computer medium of claim 122, wherein said associating first word or word string and second word or word string is enhanced by a greater frequency of occurrence of a word or word string.

10 124. The computer medium of claim 122, wherein said associating first word or word string and second word or word string is enhanced by a lesser frequency of occurrence of a word or word string.

125. The computer medium of claim 122, wherein said upper and said lower limit of said defined proximity range are equal.

15

126 The method of claim 114, further comprising:
designating either the first word or word string or the second word or word string as the first word or word string;
selecting a third word or word string, wherein said third word or word string is
20 one result from said associating step, and designating this result as the second word or word string; and
repeating said selecting, locating, defining, searching, and associating steps.

127. The computer device of claim 118, further configured to:

designating either the first word or word string or the second word or word string
as the first word or word string

selecting a third word or word string, wherein said third word or word string is
5 one result from said associating step, and designating this result as the second word or
word string; and

repeating said selecting, locating, defining, searching, and associating steps.

128. The computer medium of claim 122, further configured to:

10 designating either the first word or word string or the second word or word string
as the first word or word string

selecting a third word or word string, wherein said third word or word string is
one result from said associating step, and designating this result as the second word or
word string; and

15 repeating said selecting, locating, defining, searching, and associating steps.

129. The method of claim 105, further comprising:

designating either the first word or word string query or the second word or word
string query as the first word or word string query;

20 selecting a third word or word string, wherein said third word or word string is
one result from said associating words and word strings step, and designating this result
as the second word or word string query; and

repeating said searching, creating a second list of words and word strings, creating a second list of frequency of occurrences, creating a third list of words and word strings, and associating steps.

5 130. The computer device of claim 108, further comprising:

designating either the first word or word string query or the second word or word string query as the first word or word string query;

selecting a third word or word string, wherein said third word or word string is one result from said associating words and word strings step, and designating this result

10 as the second word or word string query; and

repeating said searching, creating a second list of words and word strings, creating a second list of frequency of occurrences, creating a third list of words and word strings, and associating steps.

15 131. The computer medium of claim 111, further comprising

designating either the first word or word string query or the second word or word string query as the first word or word string query;

selecting a third word or word string, wherein said third word or word string is one result from said associating words and word strings step, and designating this result

20 as the second word or word string query; and

repeating said searching, creating a second list of words and word strings, creating a second list of frequency of occurrences, creating a third list of words and word strings, and associating steps.

132. A method for associating words and word strings in a single language comprising:
- a. providing a collection of documents, wherein said collection includes at least one document;
 - 5 b. receiving from a user a word or word string query to be analyzed;
 - c. searching said collection of documents for the query to be analyzed and returning documents containing the query to be analyzed;
 - d. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and
 - 10 creating a Left Signature List comprising said words or word strings or both to the left of said query to be analyzed in said returned documents;
 - e. searching said collection of documents for each word and word string on said Left Signature List;
 - f. determining a user-defined amount of words or word strings or both to the right of said words or word strings or both comprising said Left Signature List and
 - 15 creating Left Anchor Lists comprising said words or word strings or both to the right of said words or word strings or both on said Left Signature List based on their frequency in a collection of documents;
 - g. determining a user-defined number of words or word strings or both to the right of said query to be analyzed in said returned documents and creating a Right Signature List comprising said words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency;

h. searching said collection of documents for each word and word string on said Right Signature List;

i. determining a user-defined number of words or word strings or both to the left of said words or word strings or both comprising said Right Signature List and
5 creating Right Anchor Lists comprising said words or word strings or both to the left of said words or word strings or both on said Right Signature List based on their frequency; and

j. ranking the results based on the frequency of each word or word string occurring on said Left Anchor Lists and the frequency of said word or word string
10 occurring on said Right Anchor Lists.

133. The method of claim 132, wherein ranking the results includes multiplying the total frequency of each word or word string occurring on said Left Anchor Lists by the total frequency of said word or word string occurring on said Right Anchor Lists.

15

134. The method of claim 132, wherein ranking the results includes adding the total frequency of each word or word string occurring on said Left Anchor Lists to the total frequency of said word or word string occurring on said Right Anchor Lists, for each word or word string occurring on at least one Left Anchor List and at least one Right
20 Anchor List.

135. The method of claim 133, wherein ranking the results is based on the total number of Left Anchor Lists and total number of Right Anchor Lists in which the word or word string occurs.

5 136. The method of claim 133, wherein ranking the results is based on user-defined parameters.

10 137. The method of claim 133, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through j. to determine and return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the list of the results of the new query.

15 138. The method of claim 133, wherein a result is modified by designating said result as a new query, and repeating steps a. through j. to determine and return the results of the new query, and modifying said result of said query based on the rank of the query on the list of the results of the new query.

20 139. The method of claim 133, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said result of said query based on the number of lists of the new queries where both the query and the result appear together.

140. The method of claim 133, wherein ranking a result is modified by designating
said each of said results as a new query, and repeating steps a. through j. to determine and
return the results of each of the new queries, and modifying said ranking of said result of
said query based on the number of lists of the new queries where both the query and the
5 result appear together.

141. The method of claim 133, wherein ranking a result is modified by designating
each of said results as a new query, and repeating steps a. through j. to determine and
return the results of each of the new queries, and modifying said ranking of the result of
10 said query based on the ranking of the query and the result on the lists of the new queries.

142. The method of claim 133, wherein a result is modified by designating each of said
results as a new query, and repeating steps a. through j. to determine and return the
results of each of the new queries, and modifying said result of said query based on the
15 ranking of the query and the result on the lists of the new queries.

143. The method of claim 133, wherein ranking a result is modified by designating
said result as a new query, repeating steps a. through i. and modifying said ranking of
said result of said query based on the words and word strings on the Left Signature List
20 and/or words and word strings on the Right Signature List of the new query that do not
appear on the Left Signature List and/or the Right Signature List of the query.

144. The method of claim 133, wherein a result is modified by designating said result
as a new query, repeating steps a. through i. and modifying said result of said query based
on the words and word strings on the Left Signature List and/or words and word strings
on the Right Signature List of the new query that do not appear on the Left Signature List
5 and/or the Right Signature List of the query.

145. The method of claim 133, wherein ranking a result is modified by designating
said result as a new query, repeating steps a. through i. and modifying said ranking of
said result of said query based on the words and word strings on the Left Signature List
10 and/or words and word strings on the Right Signature List of the query that do not appear
on the Left Signature List and/or the Right Signature List of the new query.

146. The method of claim 133, wherein a result is modified by designating said result
as a new query, repeating steps a. through i. and modifying said result of said query based
15 on the words and word strings on the Left Signature List and/or words and word strings
on the Right Signature List of the query that do not appear on the Left Signature List
and/or the Right Signature List of the new query

147. The method of claim 133, wherein a result is modified by designating said result
20 as a new query, repeating steps a. through i. and modifying said result of said query based
on the words and word strings on the Left Signature List of the query, that appear on the
Right Signature List of the new query.

148. The method of claim 133, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i. and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List of the query, that appear on the Right Signature List of the new query.

5

149. The method of claim 133, wherein a result is modified by designating said result as a new query, repeating steps a. through i. and modifying said result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.

10

150. The method of claim 133, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i. and modifying said ranking of the result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.

15

151. The method of claim 133, comprising the additional steps:

k determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a list of second word strings comprising the query and said words or word strings or both to the left of said query;

20 1. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through h.;

- m. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;
- 5 n. creating for each word string on the second list of third word strings, a second list of word and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps c. through j.;
- o. determining word strings on said list of associations that have an overlapping portion with word strings on said second list of associations; and
- 10 p. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.

152. A computer device including a processor, a memory coupled to the processor, and a program stored in the memory, wherein the computer is configured to execute the
15 program and perform the steps of:
- a. providing a collection of documents, wherein said collection includes at least one document;
- b. receiving from a user a word or word string query to be analyzed;
- c. searching said collection of documents for the query to be analyzed and
- 20 returning documents containing the query to be analyzed;
- d. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and

creating a Left Signature List comprising said words or word strings or both to the left of said query to be analyzed in said returned documents;

e. searching said collection of documents for each word and word string on said Left Signature List;

5 f. determining a user-defined amount of words or word strings or both to the right of said words or word strings or both comprising said Left Signature List and creating Left Anchor Lists comprising said words or word strings or both to the right of said words or word strings or both on said Left Signature List based on their frequency in a collection of documents;

10 g. determining a user-defined number of words or word strings or both to the right of said query to be analyzed in said returned documents and creating a Right Signature List comprising said words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency;

15 h. searching said collection of documents for words or word strings or both on said Right Signature List;

i. determining a user-defined number of words or word strings or both to the left of said words or word strings or both comprising said Right Signature List and creating Right Anchor Lists comprising said words or word strings or both to the left of said words or words strings or both on said Right Signature List based on their frequency;

20 and

j. ranking results based on the frequency of each word or word string occurring on said Left Anchor Lists and the frequency of said word or word string occurring on said Right Anchor Lists.

153. The computer device of claim 152, wherein ranking results includes multiplying the total frequency of each word or word string occurring on said Left Anchor Lists by the total frequency of said word or word string occurring on said Right Anchor Lists.

5

154. The computer device of claim 152, wherein ranking results includes adding the total frequency of each word or word string occurring on said Left Anchor Lists to the total frequency of said word or word string occurring on said Right Anchor Lists, for each word or word string occurring on one or more Left Anchor Lists and one or more

10 Right Anchor Lists.

155. The computer device of claim 152, wherein ranking results are based on the total number of Left Anchor Lists and total number of Right Anchor Lists in which the word or word string occurs.

15

156. The computer device of claim 152, wherein ranking results are based on user-defined parameters.

157. The computer device of claim 152, wherein ranking a result is modified by 20 designating said result as a new query, and repeating steps a. through j. to determine and return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the results of the new query.

158. The computer device of claim 152, wherein a result is modified by designating said result as a new query, and repeating steps a. through j. to determine and return the results of the new query, and modifying said result of said query based on the rank of the query on the results of the new query.

5

159. The computer device of claim 152, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said ranking of the result of said query based on the number of lists of the new queries where both the query and the result appear together.

10

160. The computer device of claim 152, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said result of said query based on the number of lists of the new queries where both the query and the result appear together.

15

161. The computer device of claim 152, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said ranking of the result of said query based on the ranking of the query and the return on the other new queries' lists.

20

162. The computer device of claim 152, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through i. to determine and return the results of each of the new queries, and modifying said result of said query based on the ranking of the query and the return on the other new queries' lists.

5

163. The computer device of claim 152, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the new 10 query that do not appear on the Left Signature List and/or the Right Signature List of the query.

164. The computer device of claim 152, wherein a result is modified by designating said result as a new query, repeating steps a. through g and modifying said result of said 15 query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the new query that do not appear on the Left Signature List and/or the Right Signature List of the query.

165. The computer device of claim 152, wherein ranking a result is modified by 20 designating said result as a new query, repeating steps a. through i and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the query that do not appear on the Left Signature List and/or the Right Signature List of the new query.

166. The computer device of claim 152, wherein a result is modified by designating said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Left Signature List and/or words and
5 word strings on the Right Signature List of the query that do not appear on the Left Signature List and/or the Right Signature List of the new query.
167. The computer device of claim 152, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List of the query, that appear on the Right Signature List of the new query.
10
168. The computer device of claim 152, wherein a result is modified by designating said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Left Signature List of the query, that appear on the Right Signature List of the new query.
15
169. The computer device of claim 152, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i. and modifying said ranking of the result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.
20

170. The computer device of claim 152, wherein a result is modified by designating said result as a new query, repeating steps a. through i. and modifying said result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.

5

171. The computer device of claim 152, further comprising:

k. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a list of second word strings comprising the query and said words or word strings or both to the left of said query;

10 l. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through h.;

m. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;

n. creating for each word string on the second list of said third word strings, a second list of word and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps d. through h.;

20 o. determining word strings on said list of associations that have an overlapping portion with word strings on said second list of associations; and

p. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.

172. A computer readable storage medium having stored thereon a program executable
5 by a computer processor for performing the steps of:
- a. providing a collection of documents, wherein said collection includes at least one document;
 - b. receiving from a user a word or word string query to be analyzed;
 - c. searching said collection of documents for the query to be analyzed and
10 returning documents containing the query to be analyzed;
 - d. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a Left Signature List comprising said words or word strings or both to the left of said query to be analyzed in said returned documents;
 - e. searching said collection of documents for words or word strings or both on said Left Signature List;
15 f. determining a user-defined amount of words or word strings or both to the right of said words or word strings or both comprising said Left Signature List and creating Left Anchor Lists comprising said words or word strings or both to the right of
20 said words or word strings or both on said Left Signature List based on their frequency in a collection of documents;
 - g. determining a user-defined number of words or word strings or both to the right of said query to be analyzed in said returned documents and creating a Right

Signature List comprising said words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency;

h. searching said collection of documents for words or word strings or both on said Right Signature List;

5 i. determining a user-defined number of words or word strings or both to the left of said words or word strings or both comprising said Right Signature List and creating Right Anchor Lists comprising said words or word strings or both to the left of said words or word strings or both on said Right Signature List based on their frequency; and

10 j. ranking results based on the frequency of each word or word string occurring in said Left Anchor Lists and the frequency of said word or word string occurring on said Right Anchor Lists.

173. The computer medium of claim 172, wherein ranking results includes multiplying the total frequency of each word or word string occurring on said Left Anchor Lists by the total frequency of said word or word string occurring on said Right Anchor Lists.

174. The computer medium of claim 172, wherein ranking results includes adding the total frequency of each word or word string occurring on said Left Anchor Lists to the total frequency of said word or word string occurring on said Right Anchor Lists, for each word or word string occurring on one or more Left Anchor Lists and one or more Right Anchor Lists.

175. The computer medium of claim 172, wherein ranking results are based on the total number of Left Anchor Lists and total number of Right Anchor Lists on which the word or word string occurs.

5 176. The computer medium of claim 172, wherein ranking results are based on user-defined parameters.

177. The computer medium of claim 172, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through j. to determine and
10 return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the results of the new query.

178. The computer medium of claim 172, wherein a result is modified by designating said result as a new query, and repeating steps a. through j. to determine and return the
15 results of the new query, and modifying said result of said query based on the rank of the query on the results of the new query.

179. The computer medium of claim 172, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through j. to
20 determine and return the results of each of the new queries, and modifying said ranking of the result of said query based on the number of lists of the new queries where both the query and the result appear together.

180. The computer medium of claim 172, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said result of said query based on the number of lists of the new queries where both the query and the result
5 appear together.

181. The computer medium of claim 172, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and return the results of each of the new queries, and modifying said ranking
10 of the result of said query based on the ranking of the query and the return on the lists of the new queries on which they both appear.

182. The computer medium of claim 172, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through j. to determine and
15 return the results of each of the new queries, and modifying said result of said query based on the ranking of the query and the return on the lists of the new queries on which they both appear .

183. The computer medium of claim 172, wherein ranking a result is modified by
20 designating said result as a new query, repeating steps a. through i and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the new

query that do not appear on the Left Signature List and/or the Right Signature List of the query.

184. The computer medium of claim 172, wherein a result is modified by designating
5 said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the new query that do not appear on the Left Signature List and/or the Right Signature List of the query.

10 185. The computer medium of claim 172, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i and modifying said ranking of the result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the query that do not appear on the Left Signature List and/or the Right Signature List of the new query.

15 186. The computer medium of claim 172, wherein a result is modified by designating said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Left Signature List and/or words and word strings on the Right Signature List of the query that do not appear on the Left Signature List and/or the Right Signature List of the new query.

20

187. The computer medium of claim 172, wherein ranking a result is modified by designating said result as a new query, repeating steps a. through i and modifying said

ranking of the result of said query based on the words and word strings on the Left Signature List of the query, that appear on the Right Signature List of the new query.

188. The computer medium of claim 172, wherein a result is modified by designating
5 said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Left Signature List of the query, that appear on the Right Signature List of the new query.

189. The computer medium of claim 172, wherein ranking a result is modified by
10 designating said result as a new query, repeating steps a. through i. and modifying said ranking of the result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.

190. The computer medium of claim 172, wherein a result is modified by designating
15 said result as a new query, repeating steps a. through i and modifying said result of said query based on the words and word strings on the Right Signature List of the query, that appear on the Left Signature List of the new query.

191. The computer medium of claim 172, further comprising:
20 k. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a list of second word strings comprising the query and said words or word strings or both to the left of said query;

- l. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through h.;
 - m. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;
 - n. creating for each word string on the second list of said third word strings, a second list of word and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps d. through h.;
 - o. determining word strings on said list of associations that have an overlapping portion with word strings on said second list of associations; and
 - p. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.

- 15
192. A method for associating words and word strings in a language comprising:
 - a. providing a collection of documents, wherein said collection includes at least one document;
 - b. receiving from a user a word or word string query to be analyzed;
 - c. searching said collection of documents for the query to be analyzed and returning documents containing the query to be analyzed;

- d. determining a user-defined number of words or word strings of user-defined size or both to the left and right of the query in said returned documents containing the query to be analyzed;
- e. returning a list with an entry or plurality of entries, wherein said entry or 5 said plurality of entries contain said determined words or word strings or both to the left and right of the query in said returned documents;
- f. searching said collection of documents for said entry or plurality of entries in said returned list; and
- g. returning a list of words or word strings of user defined size or both that 10 occur most frequently between said determined words or word strings or both to the left and right of said query in said returned documents.

- 193. The method of claim 192, wherein said returned list of words or word strings or both is ranked based on the number of unique said determined words or word strings or 15 both to the left and right of said returned list of words.
- 194. The method of claim 192 or 193, wherein said returned list of words or word strings or both is ranked based on user-defined parameters.
- 20 195. The method of claim 192, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through g. to determine and return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the results of the new query.

196. The method of claim 192, wherein a result is modified by designating said result
as a new query, and repeating steps a. through g. to determine and return the results of the
new query, and modifying said result of said query based on the rank of the query on the
5 results of the new query.

197. The method of claim 192, wherein ranking a result is modified by designating
each of said results as a new query, and repeating steps a. through g. to determine and
return the results of the new queries, and modifying said ranking of the result of said
10 query based on the number of lists of the new queries where both the query and the result
appear together.

198. The method of claim 192, wherein a result is modified by designating each of said
results as a new query, and repeating steps a. through g. to determine and return the
15 results of the new queries, and modifying said result of said query based on the number
of lists of the new queries where both the query and the result appear together.

199. The method of claim 192, wherein ranking a result is modified by designating
each of said results as a new query, and repeating steps a. through g. to determine and
20 return the results of the new queries, and modifying said ranking of the result of said
query based on the ranking of the query and the result on the lists of the new queries
where both the query and the result appear together.

200. The method of claim 192, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said result of said query based on the ranking of the query and the result on the lists of the new queries where both the query and the
5 result appear together.

201. The method of claim 192, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left of the new query and to the right of the new
10 query, and modifying said ranking of the result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear to the left and/or to the right of the new query.

202. The method of claim 192, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left of the new query and to the right of the new query, and modifying said result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear to the left and/or to the right of the new query.
15
20

203. The method of claim 192, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left of the new query and to the right of the new

query, and modifying said ranking of the result of said query based on the words or word strings or both to the left of the new query and/or words or word strings or both to the right of the new query that do not appear to the left and/or to the right of the query.

5 204. The method of claim 192, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left of the new query and to the right of the new query, and modifying said result of said query based on the words or word strings or both to the left of the new query and/or words or word strings or both to the right of the new query that

10 do not appear to the left and/or to the right of the query.

205. The method of claim 192, further comprising:

h. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and

15 creating a list of second word strings comprising the query and said words or word strings or both to the left of said query;

i. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through g.;

20 j. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;

- k. creating for each word string on the second list of third word strings, a second list of word and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps c. through g.;
- l. determining word strings on said list of associations that have an overlapping portion with a word string on said second list of associations; and
- 5 m. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.

206. A computer device including a processor, a memory coupled to the processor, and
10 a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:
- a. providing a collection of documents, wherein said collection includes at least one document;
- b. receiving from a user a word or word string query to be analyzed;
- 15 c. searching said collection of documents for the query to be analyzed and returning documents containing the query to be analyzed;
- d. determining a user-defined number of words or word strings of user-defined size or both to the left and right of the query in said returned documents containing the query to be analyzed;
- 20 e. returning a list with an entry or plurality of entries, wherein said entry or said plurality of entries contain said determined words or word strings or both to the left and right of the query in said returned documents;

- f. searching said collection of documents for said entry or plurality of entries in said returned list; and
- g. returning a list of words or word strings of user defined size or both that occur most frequently between said determined words or word strings or both to the left 5 and right of said query in said returned documents.

207. The computer device of claim 206, wherein said returned list of words or word strings or both is ranked based on the number of unique said determined words or word strings to the left and right of said returned list of words.

10

208. The computer device of claim 206 or 207, wherein said returned list of words or word strings or both is ranked based on user-defined parameters.

209. The computer device of claim 206, wherein ranking a result is modified by 15 designating said result as a new query, and repeating steps a. through g. to determine and return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the results of the new query.

210. The computer device of claim 206, wherein a result is modified by designating 20 said result as a new query, and repeating steps a. through g. to determine and return the results of the new query, and modifying said result of said query based on the rank of the query on the results of the new query.

211. The computer device of claim 206 wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said ranking of the result of said query based on the number of lists of the new queries where both the query
5 and the result appear together.

212. The computer device of claim 206, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said result of said query based on the
10 number of lists of the new queries where both the query and the result appear together.

213. The computer device of claim 206, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said ranking of the result of said query based on the ranking of the query and the result on the lists of the
15 new queries that they both appear on.

214. The computer device of claim 206, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said result of said query based on the
20 ranking of the query and the result on the lists of the new queries that they both appear on.

215. The computer device of claim 206, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left and/or to the right of the new query, and ranking the result of said query based on the words or word strings or both to the left of
5 the new query and/or words or word strings or both to the right of the new query that do not appear to the left and/or right of the query.

216. The computer device of claim 206, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return
10 words or word strings or both to the left and/or to the right of the new query, and modifying said result of said query based on the words or word strings or both to the left of the new query and/or words or word strings or both to the right of the new query that do not appear to the left and/or right of the query.

15 217. The computer device of claim 206, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return words or word strings or both to the left and/or to the right of the new query, and ranking the result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear
20 to the left and/or right of the new query.

218. The computer device of claim 206, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return

words or word strings or both to the left and/or to the right of the new query, and modifying said result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear to the left and/or right of the new query.

5

219 The computer device of claim 206, further comprising:

h. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a list of second word strings comprising the query and said words or word

10 strings or both to the left of said query;

i. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through g.;

j. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;

k. creating for each word string on the second list of third word strings, a second list of words and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps c. through g.;

20 l. determining word strings on said list of associations that have an overlapping portions with a word string on said second list of associations; and

- m. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.
220. A computer readable storage medium having stored thereon a program executable
5 by a computer processor for performing the steps of:
- a. providing a collection of documents, wherein said collection includes at least one document;
 - b. receiving from a user a word or word string query to be analyzed;
 - c. searching said collection of documents for the query to be analyzed and
10 returning documents containing the query to be analyzed;
 - d. determining a user-defined number of words or word strings of user-defined size or both to the left and right of the query in said returned documents containing the query to be analyzed;
 - e. returning a list with an entry or plurality of entries, wherein said entry or
15 said plurality of entries contain said determined words or word strings or both to the left and right of the query in said returned documents;
 - f. searching said collection of documents for said entry or plurality of entries in said returned list; and
 - g. returning a list of words or word strings of user defined size or both that
20 occur most frequently between said determined words or word strings or both to the left and right of said query in said returned documents.

221. The computer medium of claim 220, wherein said returned list of words or word strings or both is ranked based on the number of unique said determined words or word strings to the left and right of said query on said returned list of words.

5 222. The computer medium of claim 220 or 221, wherein said returned list of words or word strings or both is ranked based on user-defined parameters.

10 223. The computer medium of claim 220, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through g. to determine and return the results of the new query, and modifying said ranking of the result of said query based on the rank of the query on the results of the new query.

15 224. The computer medium of claim 220, wherein a result is modified by designating said result as a new query, and repeating steps a. through g. to determine and return the results of the new query, and modifying said result of said query based on the rank of the query on the results of the new query.

20 225. The computer medium of claim 220 wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of each of the new queries, and modifying said ranking of the result of said query based on the number of lists of the new queries where both the query and the result appear together.

226. The computer medium of claim 220, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said result of said query based on the number of lists of the new queries where both the query and the result appear together.

5

227. The computer medium of claim 220, wherein ranking a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said ranking of the result of said query based on the ranking of the query and the result on the lists of the new queries that they both appear on.

10

228. The computer medium of claim 220, wherein a result is modified by designating each of said results as a new query, and repeating steps a. through g. to determine and return the results of the new queries, and modifying said result of said query based on the ranking of the query and the result on the lists of the new queries that they both appear on.

15

229. The computer medium of claim 220, wherein ranking a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return the words or word strings or both to the left and to the right of the new query, and modifying said ranking of the result of said query based on the words or word strings or both to the left of the new query and/or words or word strings or both to the right of the new query that do not appear to the left and/or right of the query.

20

230. The computer medium of claim 220, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return the words or word strings or both to the left and to the right of the new query, and modifying
5 said result of said query based on the words or word strings or both to the left of the new query and/or words or word strings or both to the right of the new query that do not appear to the left and/or right of the query.

231. The computer medium of claim 220, wherein ranking a result is modified by
10 designating said result as a new query, and repeating steps a. through e. to determine and return the words or word strings or both to the left and to the right of the new query, and modifying said ranking of the result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear to the left and/or right of the new query.

15

232. The computer medium of claim 220, wherein a result is modified by designating said result as a new query, and repeating steps a. through e. to determine and return the words or word strings or both to the left and to the right of the new query, and modifying
20 said result of said query based on the words or word strings or both to the left of the query and/or words or word strings or both to the right of the query that do not appear to the left and/or right of the new query.

233. The computer medium of claim 220, further comprising:

- h. determining a user-defined amount of words or word strings or both to the left of said query to be analyzed in said returned documents based on their frequency and creating a list of second word strings comprising the query and said words or word strings or both to the left of said query;
- 5 i. creating for each word string on the list of second word strings, a list of word and word string associations by designating each word string on the list of second word strings as a new query and repeating steps c. through g.;
- j. determining a user-defined amount of words or word strings or both to the right of said query to be analyzed in said returned documents based on their frequency
- 10 and creating a second list of third word strings comprising the query and said words or word strings or both to the right of said query;
- k. creating for each word string on the second list of third word strings, a second list of words and word string associations by designating each word string on the second list of third word strings as a new query and repeating steps c. through g.;
- 15 l. determining word strings on said list of associations that have an overlapping portions with a word string on said second list of associations; and
- m. identifying the word or word strings in the overlapping portions of the overlapping word strings as synonyms or near synonyms of the query.
- 20 234. A method for content conversion within a single language comprising the following steps:
- a. providing a first plurality of word strings;

- b. providing a second plurality of word strings, wherein each of said word strings in said second plurality corresponds to one of said word strings in said first plurality in a synonymous or near synonymous manner;
- c. receiving a word string query to be analyzed;
- 5 d. parsing said word string query into plurality of subset word strings, wherein a portion of each subset overlaps with a second portion of its adjoining subset or subsets;
- e. analyzing each of said parsed subset word strings to identify, using said second plurality of word strings, synonymous word strings for each of said parsed subset word strings; and
- 10 f. replacing any parsed subset word string with a synonymous word string where it overlaps with said adjoining subsets.

235. A computer device including a processor, a memory coupled to the processor, and
- 15 a program stored in the memory, wherein the computer is configured to execute the program and perform the steps of:
 - a. providing a first plurality of word strings;
 - b. providing a second plurality of word strings, wherein each of said word strings in said second plurality corresponds to one of said word strings in said first plurality in a synonymous or near synonymous manner;
 - 20 c. receiving a word string query to be analyzed;

- d. parsing said word string query into plurality of subset word strings, wherein a portion of each subset overlaps with a second portion of its adjoining subset or subsets;
 - e. analyzing each of said parsed subset word strings to identify, using said second plurality of word strings, synonymous word strings for each of said parsed subset word strings; and
 - f. replacing any parsed subset word string with a synonymous word string where it overlaps with said adjoining subsets.
- 10 236. A computer readable storage medium having stored thereon a program executable by a computer processor for performing the steps of:
- a. providing a first plurality of word strings;
 - b. providing a second plurality of word strings, wherein each of said word strings in said second plurality corresponds to one of said word strings in said first plurality in a synonymous or near synonymous manner;
 - c. receiving a word string query to be analyzed;
 - d. parsing said word string query into plurality of subset word strings, wherein a portion of each subset overlaps with second portion of its adjoining subset or subsets;
 - e. analyzing each of said parsed subset word strings to identify, using said second plurality of word strings, synonymous word strings for each of said parsed subset word strings; and

f. replacing any parsed subset word string with a synonymous word string where it overlaps with said adjoining subsets.